1997-1

ON THE USE OF INJURY SCORES FOR JUDGING THE ACCEPTABILITY OF RESTRAINING TRAPS

RICHARD M. ENGEMAN
National Wildlife Research Center, 1716 Heath Parkway, Fort Collins, CO 80524–2719, USA

HEATHER W. KRUPA
National Wildlife Research Center, 1716 Heath Parkway, Fort Collins, CO 80524–2719, USA

JOHN KERN
Statistics Department, University of Wyoming, Laramie, WY 82071–3332, USA

Abstract: Injury scores have been used as numerical tallies describing injury levels of animals caught in restraining traps. Using data on the severity of injuries inflicted on the limbs of 47 foxes, we evaluated injury scores as a tool for judging the acceptability of trap type. Data from 6 veterinarians indicated moderate to serious potential for interobserver variation, and the potential for inconsistencies between scored assessments and the observers' general perceptions of acceptability of the injury level. Although these issues may be resolved through training, we also discuss conceptual problems with the quantitative properties of any scoring system and suggest quantifying the injuries more directly than through the use of scores.

Key words: injury category, observer consistency, quantitative properties.

J. Wildl. Res. 2(2): 124–127, 1997

INTRODUCTION

In recent years there has been considerable and increasing public pressure to reduce the level of injury and perceived suffering in the capture of wild furbearing animals. As such, "humaneness" of restraining traps is the criterion for international standards being developed for application to commerce in animal furs obtained through trapping. Nonetheless, both the perceptions of the amount of suffering incurred by a trapped animal and the amounts that should be morally tolerated are highly-variable from observer to observer, and from culture to culture.

Injury scores often have been used to quantify the extent of injury incurred by a trapped animal and a number of studies have used injury scores to compare the severity of injuries sustained by animals caught in different types of leg-hold traps (e.g., Tullar 1984, Olsen et al. 1986, Linhart et al. 1988, Olsen et al. 1988, Onderka et al. 1990, Phillips et al. 1992). Here, we address the issue of using injury scores as a measure for trap acceptability.

Injury scores are numerical tallies assigned by an observer to the injuries of a captured animal. A scoring system may address only injuries to the limb held by the trap, or also it may include injuries sustained to other areas (such as the mouth). Potential injuries are assigned a point value based on consideration such as pain, severity of the injury and its potential for recovery, and whether the animal could reasonably be released (if desired). When an animal is examined, the point values assigned to each injury are summed for all injuries to that animal. Different authors have used different injury categories and scores in their systems, but the general concepts are similar.

As an example, the scoring system used by Olsen et al. (1986) is reproduced in Table 1. The original application of that scoring system was to compare 2 trap types (using 1 observer) — not to provide a measure of acceptability for a trap type. However, we use it as an example because it is one of the most-cited references concerning the use of injury scores.

In the present paper, we examine the use of injury scores for the purpose of trap evaluations in the following contexts: (1) the consistency among observers for scoring injuries to limbs of trapped animals, (2) the observers' perceived "acceptability" of injuries as compared to "acceptability" based on a pre-defined threshold for the injury score, (3) the quantitative properties of a scoring system, (4) the questions to which the scoring system is applied, (5) alternate quantitative approaches for addressing the same questions.

The final 3 objectives are conceptual in nature, and are discussed as such in quantitative/statistical terms. The first 2 objectives are also conceptual in that they address questions that should be asked of any scoring system used to evaluate injuries to trapped animals. However, we address the first 2 objectives by analyzing data from veterinarian's evaluations of limbs from trapped animals.

We are grateful to Gordon Batcheler and Kathy Loconti-Lee for providing the New York State Department of Environmental Conservation data and for reviewing this manuscript. We thank the veterinarians who participated in the scoring exercises for their efforts. Frederick F. Gilbert was also supportive in the use of those data and reviewed the manuscript. Reviews were also provided by Albert Dale, James Miles, Michael W. Fall and Robert L. Phillips.

Table 1, Injury scoring system for foothold trap-induced injuries to coyotes reproducted from Olsen et al. (1986).

Description of injury	Points scored
Apparently normal	0
Edematous swelling and hemorrhage	5
Cutaneous laceration <2 cm	5
Cutaneous laceration >2 cm	10
Tendon and ligament laceration	20
Joint subluxation	30
Joint luxation	50
Compression fracture above or below carpus or tarsus	30
Simple fracture at or below carpus or tarsus	50
Compound fracture at or below carpus or tarsus	75
Simple fracture above carpus or tarsus	100
Compound fracture above carpus or tarsus	200
Amputation	400

MATERIALS AND METHODS

Injury data were derived from trapped fox limbs (n=47) collected by the New York State Department of Environmental Conservation in 1990. The injury score results were provided to us after examination of the limbs was conducted by a group of 6 experts in veterinary pathology (A, B, C, D, F) from several countries. Each observer was highly experienced with trap injuries and evaluated the injuries to the limbs by 2 methods: (1) application of an injury score (Olsen et al. 1986), and (2) the observer's perception of the severity of injury to the lamb. A work station was set up for each of the 6 observers where a limb was randomly selected and presented for dissection and scoring. The observers then rotated to the next work station and scored the dissected limb at that station. In this manner each observer scored each limb, with the first observer of each limb performing the dissection. At the time of scoring, the observers were not aware of previous scores on each limb. After rotating back to their original work stations, a new set of 6 limbs was presented for dissection and the process was repeated. The other evaluation method was a yes-no judgment for each limb as to whether an injury, or combination of injuries, was "acceptable", or not, based on each observer's individual perception of humaneness.

Observer consistency for score magnitudes was compared using Friedman's 2-way layout (e.g. Hollander and Wolfe 1971). To evaluate the effect in the practical sense of how any differences in score magnitudes among observers relate to the "acceptability" of an injury, we used the threshold score of 75 to define the limit of acceptability (this value, used in a conceptual context here, was based on the general consensus from communications with the members of the International Organization for Standardization (ISO) Technical Committee 191 working on the development of trap standards). However, this particular value is not as important as the implications from the results. These (dichotomized acceptable or unacceptable) data thus derived from each observer's score on each limb were analyzed using Cochran's Q test (e.g., Winer 1971), which basically is a repeated-measures analysis for binary data to compare rejection rates among the observers. In addition to differences in limb scores and rejection rates among observers, we also calculated correlations among the observers to ascertain how well their scores paralleled each other across all of the limbs. We also were interested to see how well the perceived acceptability of an injury compared to the acceptability based on an injury threshold. In addition to whether or not the injury score exceeded 75 for each limb, each observer produced a second binary appraisal of acceptability based on his overall impression or perception of the injury level. For each observer, we used McNemar's test (e.g. Sokal and Rohlf 1981) to discern whether the proportion of limbs judged as unacceptable was the same using the 2 appraisal methods. Kendall's tau (Agresti 1990), was used to examine the strength of the relationship between the 2 appraisals of acceptability.

RESULTS

Interobserver variation

Friedman's test indicated a strong difference (p<0.0001) in score magnitudes among observers. The largest differences were between observer A and all others, and between observer E and 4 of the others (mean rank for observers A through F were, respectively: 4.5, 3.6, 3.5, 2.7, 3.4). Cochran's Q test showed a strong difference among observers (p<0.005) for the proportion of limbs scored as unacceptable (scores >75). Corresponding to the above comparison of score magnitudes, observer E had the lowest rejection rate (11%) based on a threshold value of 75 for rejection.

However, in contrast to the score-magnitude results, the highest rejection rates were produced by observers B and D (21%), rather than observer A (15%). Table 2 presents the correlation matrix among the 6 observers. A number of the correlations are quite high, but observer F generally correlates only moderately with the others (4 of 5 correlations), and observer A does not have particularly-high correlations (range 0.46–0.91).

Score acceptability versus perceived acceptability

A difference was detected using McNemar's test (range, p=0.005, to p=0.046) between the proportion perceived as unacceptable and the proportion scored as unacceptable for 4 of the 6 observers (A, B, C, E) and one of the other two observers

Table 2. Correlations among 6 observers using injury scores to evaluate fox limbs with trapping injuries.

Observer	B	C	D	E	F
A	0.91	0.85	0.76	0.86	0.46
B	—	0.96	0.91	0.97	0.45
C	—	—	0.94	0.96	0.41
D	—	—	—	0.93	0.88
E	—	—	—	—	0.45

(D) had a borderline difference (p=0.083) between the 2 methods. Observers A through E each perceived a higher proportion of limbs to have unacceptable injuries than was indicated by the scoring system. Only for observer F was a difference not apparent. A particularly-strong relationship was not consistently indicated between the appraisal methods. Observer D had the highest tau of 0.84, whereas half of the observers had a tau below 0.70, with observer E being lowest at 0.40.

Evaluation of scoring system properties

A scoring system is not a direct measure of the severity of injuries. For example, a limb with "edematous swelling and hemorrhage" plus a "cutaneous laceration" cm" plus "tendon a ligament laceration" would score as 30 (Olsen et al. 1986). This would be considered the identical level of injury severity as a "compression fracture above or below the carpus or tarsus" — but there is no way to know or measure this. Similarly, scores do not provide a direct measure of relative differences in severity of injury between limbs. That is, it can only be a presumption that the magnitude of 10 increase in severity from "cutaneous laceration 2 cm" to "tendon and ligament laceration" would be equivalent to the magnitude 10 increase in severity from "tendon and ligament laceration" to "joint subluxation". Values for injury categories are somewhat arbitrary numerical assignments, and injury scores are abstractions of the severity of the combined injuries to an animal with many possibilities for obtaining a particular score. Therefore, it is inappropriate to treat them as one-to-one measurements of the severity of injuries.

DISCUSSION

Application of a scoring system

Although we do not promote the use of a scoring system in this paper, for comparative purposes we consider how one could be used. We assume that a scoring system would be intended for application to a sample of an animal species caught in a particular type of restraining trap to judge whether or not that trap type is suitable for the capture of animals. If a scoring system is to be applied for this purpose, a logical (and undoubtedly simplified) scenario for its development and implementation might involve the following sequence: (1) Define the injury categories to be considered; (2) Decide on a value to be applied to each category, (3) Decide how to combine these values into a score; (4) Decide on a threshold at which an injury score becomes unacceptable; (5) Decide on the minimum number of captures that would provide an adequate test of a trap type; (6) Decide on the frequency of occurrence of an unacceptable score in a sample of animals that would disqualify the use of a trap. This approach would require subjective decisions on many levels — each increasingly abstract.

Complicating this picture is the existence of a variety of scoring systems already applied to trap injuries, with a number of these reported in the literature. Although the system by Olsen et al. (1986) formed the basis for the development of the other systems, modifications have made a comparison of scores from one system to the next uncertain at best. For example, the system produced by Onderka et al. (1990); Table 3, gives a more thorough accounting of injury types than the Olsem system. The same injury scored under the two systems could produce considerably different scores — thus making it difficult to contrast and compare severity of injuries. As an extreme example, an amputation of a digit would score 400 in the Olsen system, but depending on the nature of the amputation, could score as little as 30 in the Onderka system.

Table 3. Injury scoring system for foothold trap-induced injuries to coyotes reproducted from Onderka et al. (1990).

Description of injury	Points scored
Apparently normal	0
Edematous swelling and/or hemorrhage	1–5
Cutaneous laceration <2 cm	5
Cutaneous laceration >2 cm	10
Subcutaneous muscle laceration or maceration	10 20
Tendon or ligamenta maceration with partial severance	20–40
Partial fracture of metacarpi or metatarsi	30
Fracture of digits	30–40
Amputation of digits	30–40
Joint luxation of digits	50
Simple fracture below carpus or tarsus	50
Severance of tendons below carpus or tarsus	50
Compound fracture below carpus or tarsus	75
Simple fracture above carpus or tarsus	100
Compound fracture above carpus or tarsus	200
Luxated elbow or hock joint	200–300
Amputation of limb	400

Alternative approaches

We suggest more-directly quantifying the issue of whether a trap type causes unacceptable injury to the animals captured. As with a scoring system, decisions would be required to define the injury categories to be considered and which injuries are unacceptable. If a particular injury is deemed unacceptable, then the frequency with which that injury occurs would be a logical criterion for determining the acceptability of a trap type. More generally, the frequency of unacceptable injuries in aggregate could be used as the criterion. The observing veterinarian should be permitted the flexibility to include severity, or number, of injury type into determination of injury acceptability (note that to some degree this is attempted in the Onderka system in Table 3, where some injuries are allowed a range of scores). As an example, a laceration may be acceptable unless an excess amount of dirt had been ground into the wound.

An acceptable catch would be defined as any capture resulting in no unacceptable injuries. As with a scoring system, decisions would be required as to the minimum number of captures that would yield an adequate test of a trap and the threshold percentage of acceptable captures. We consider 2 possibilities for implementation of such a system.

First, decisions could be made regarding the frequency of occurrence for each injury type that would define a trap as unacceptable. Any occurrence of some injuries, such as compound fractures, may disqualify a trap type, whereas a frequency of 1 in 10 or 20 might be deemed acceptable for a less severe injury. This system allows for a great deal of flexibility in defining how trap types may be determined as acceptable or not, but, also requires decisions at a number of stages.

Another more direct approach for relating injuries to trap acceptability would involve grouping the injuries types as negligible, moderate, or severe. After deciding upon a minimum sample size, a frequency of occurrence could be defined for each injury level grouping whereby a trap type would be disqualified. This system requires fewer decisions at fewer steps and appears easier to implement into practice.

CONCLUSIONS

Injury scores are not direct measurements of injury levels. Inconsistent scores among observers can result, as well as inconsistencies between the observers perception of injury level and the threshold defined for the score. We believe that training and experience with a particular scoring system could alleviate repeatability problems, but our results still warn of the potential for problems in repeatability. Application of a scoring system requires decisions on several levels of increasing abstraction from the physical injuries. We feel that trap assessments would be more straight-forward if they were based on decisions as to what injuries should be considered unacceptable and how frequently they can occur before a trap type is deemed unacceptable.

LITERATURE CITED

AGRESTI, A. 1990. Categorical Data Analysis. John Wiley & Sons, New York. 558 pp.

HOLLANDER, M, AND D.A. WOLFE. 1973. Nonparametric Statistical Methods. John Wiley & Sons, New York. 503 pp.

LINHART, S.B., F.S. BLOM, G.J. DASCH, R.M. ENGEMAN, AND G.H. OLSEN. 1988. Field evaluation of padded jaw coyote traps. Proc. Vert. Pest Conf. 13: 226–229.

OLSEN, G.H., S.B. LINHART, R.A. HOLMES, G.J. AND C.B. MALE. 1986. Injuries to coyotes caught in padded and unpadded steel foothold traps. Wildl. Soc. Bull. 14: 219–223.

——, R.G. LINSCOMBE, V.L. WRIGHT, AND R.A. HOLMES. 1988. Reducing injuries to terrestrial furbearers by using padded foothold traps. Wildl. Soc. Bull. 16: 303–307.

ONDERKA, D.K., D.L. SKINNER, AND A.W. TODD. 1990. Injuries to coyotes and other species caused by four models of footholding devices. Wildl. Soc. Bull. 16: 175–182.

PHILLIPS, R.L., F.S. BLOM, G.J., AND J.W. GUTHRIE. 1992. Field evaluation of three types of coyote traps. Proc. Vert. Pest Conf. 15: 393–395.

SOKAL, R.R., AND F.J. ROHLF. 1981. Biometry. W.H. Freeman and Co., New York. 859 pp.

TULLAR, B.F. 1984. Evaluation of padded leghold trap for capturing foxes and raccoons. New York Fish and Game J. 31: 97–103.

WINER, B.J. 1971. Statistical Principles in Experimental Design. McGraw-Hill, New York. 907 pp.